

(12) PATENTIJULKAIKU
PATENTSKRIFT

(10) FI 103156 B

(45) Patentti myönnetty - Patent beviljats

30.04.1999

(51) Kv.1k.6 - Int.kl.6

G 06F 17/28

(21) Patentihakemus - Patentansökaning

964742

(22) Hakemispäivä - Ansökningsdag.

28.11.1996

(24) Alkupäivä - Löpdag

28.11.1996

(41) Tullut julkiseksi - Blivit offentlig

29.05.1998

SUOMI-FINLAND
(FI)Patentti- ja rekisterihallitus
Patent- och registerstyrelsen

(73) Haltija - Innehavare

1. Becks, Ari, Osuuskunnantie 100 C 3, 00660 Helsinki, (FI)
2. Heikkilä, Simo Sakari, Upseerinkatu 1-5 A 2, 15700 Lahti, (FI)

(72) Keksijä - Uppfinner

1. Becks, Ari, Osuuskunnantie 100 C 3, 00660 Helsinki, (FI)
2. Heikkilä, Simo Sakari, Upseerinkatu 1-5 A 2, 15700 Lahti, (FI)

(74) Asiamies - Ombud: Berggren Oy Ab, Jaakonkatu 3 A, 00100 Helsinki

(54) Keksinnön nimitys - Uppfinningens benämning

Menetelmä ja järjestely informaation kääntämiseksi
Förvarande och anordning för att översätta information

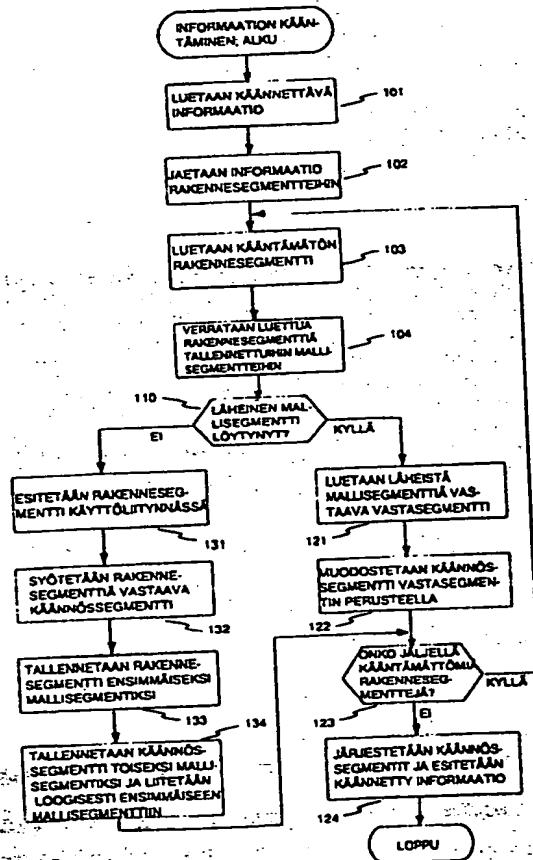
(56) Viitejulkaisut - Anfördta publikationer

EP A 0672989 (G 06F 17/28), US A 5267156 (G 06F 15/38).
Proceedings of International Join Conference of Artificial Intelligence, 28 Aug.- 3 Sept. 1993
Chambery, France; Morgan Kaufman Publishers, USA, Part vol. 2, 1993; H. Kitano: "A comprehensive and
practical Model of memory-based machine translation", p. 1276-82, Artificial Intelligence,
vol. 75, nro 1, 1995, Netherlands, S. Sato: "MBT2:a method of combining fragments of examples in
example-based translation", p. 31-49, Advances in Case-Based Reasoning, Third European Workshop,
EWCBR-96, Proceedings, Springer-Verlag, Berlin, Germany, B. Collins et al.: "Adaptation-guided
retrieval in EBM: a case-based approach to machine translation", p. 91-104:1,2,10,11,
1995 IEEE International Conference on Systems, Man and Cybernetics, Intelligent Systems for
the 21st Century, Part vol. 2, IEEE, New York, USA, R. Jain et al.: "Role of examples in translation",
p. 1615-20:1,2,10,11, Proceedings of the AAAI Fall Symposium Series, 23-25 Oct. 1992, Cambridge,
MA, USA, AAAI Press, 1993, Menlo Park, CA USA, D. Jones: "Virtual machine translation",
p. 84-88:1,2,10,11, IEICE Transaction on Information and Systems, vol. E75-D, nro 4, 1992, Japan,
E. Sumita et al.: "Example-based transfer of Japanese adnominal particles into English",
p. 585-594:1,2,10,11

(57) Tiivistelmä - Sammandrag

Keksinnön kohteena on menetelmä ja järjestely ensimmäisen kielen merkkijonona esitetyn informaation kääntämiseksi toisen kielen merkkijonoksi. Keksintöä sovelletaan edullisesti teksti-informaation kääntämiseen tietokoneessa. Keksinnön eräään oivalluksena on se, että käännitteivä informaatio jaetaan rakennesegmentteihin (102) ja suoritetaan käännös rakennesegmenttikohtaisesti (122). Käännös suoritetaan tietämyskantaan tallennettujen mallisegmenttien ja sääntöjen perusteella. Tietämyskannan sisältämää tietoa laajennetaan edullisesti siten, että käännöksen suoritukseen liittyen käyttäjä syöttää tarvittaessa käyttöliittynän välityksellä uusien rakennesegmenttien käännöksiä (132), jotka sitten tallennetaan mallisegmentteinä tietämyskantaan (133, 134). Keksinnön mukaisen ratkaisun ansiosta käännöksiä suorittavassa laitteistossa riittää pienempi muistikapasiteetti ja prosessorin nopeus. Lisäksi ohjelointityötä tarvitaan huomattavasti vähemmän ja laitteiston toimintaa voidaan kehittää ilman ohjelinistopäivityksiä.

Uppfinningen avser ett förfarande och arrangemang för översättning av information presenterad som en teckensträng på ett första språk till en teckensträng på ett andra språk. Uppfinningen tillämpas fördelaktigt vid översättning av textinformation i dator. En av idéerna med uppsättningen är att informationen som skall översättas indelas i struktursegment (102) varefter översättningen utsörs segmentvis (122). Översättningen utsörs utgående från modellsegment och regler som lagrats i en kunskapsbas. Den information som ingår i kunskapsbasen expanderas fördelaktigt så, att användaren i samband med utförande av översättningen vid behov via förmedling av användargränsnivå matar in översättningar (132) av nya struktursegment, vilka översättningar sedan lagras i kunskapsbasen (133, 134) som modellsegment. Tack vare utförandet enligt uppsättningen är en mindre minneskapacitet och snabbhet hos processorn tillräckliga hos apparaturen som utför översättandet. Dessutom krävs betydligt mindre programmeringsarbete samtidigt som funktionen hos apparaturen kan utvecklas utan uppdatering av programvaran.



Menetelmä ja järjestely informaation käänämiseksi - Förfarande och anordning för att översätta information

5 Keksinnön kohteena on menetelmä ja järjestely ensimmäisen kielen merkkijonona esitetyn informaation käänämiseksi toisen kielen merkkijonoksi. Keksintöä sovelletaan edullisesti teksti-informaation käänämiseen tietokoneessa.

10 Ennestään tunnetaan lingvistiikkaan perustuvia menetelmiä teksti-informaation käänämiseksi tietokoneella. Tällaisia menetelmiä on esitetty mm. seuraavissa julkaisuissa:

[1] EP-A-0 672 989,
 [2] US-A-5 267 156,
 15 [3] Proceedings of International Joint Conference of Artificial Intelligence, 28 Aug.-3. Sept. 1993, Chambery, France; Morgan Kaufman Publishers, USA, Part vol. 2, 1993, H. Kitano: "A comprehensive and practical model of Memory-based machine translation", s. 1276-82,
 [4] Artificial Intelligence, vol. 75, no 1, 1995, Netherlands, S. Sato.: "MBT2: a method of combining fragments of examples in example-based translation", s. 31-49,
 20 [5] Advances in Case-Based Reasoning, Third European Workshop, EWCBR-96, Proceedings, Springer-Verlag, Berlin, Germany, B. Collins et al.: "Adaption-guided retrieval in EBMT: a case-based approach to machine translation", s. 91-104,
 [6] 1995 IEEE International Conference on Systems, Man and Cybernetics, Intelligent Systems for the 21st Century, Part vol. 2, IEEE, New York, USA, R. Jain et al.: "Role of examples in translation", s. 1615-20,
 25 [7] Proceedings of the AAAI Fall Symposium Series, 23-25 Oct. 1992, Cambridge, MA, USA, AAAI Press, 1993, Menlo Park, CA USA, D. Jones: "Virtual machine translation", s. 84-88 ja
 [8] IEICE Transaction on Information and Systems, vol. E75-D, no. 4, 1992, Japan, E. Sumita et al.: "Example-based transfer of Japanese adnominal particles into English", s. 585-594.

30 35 Näissä ennestään tunnetuissa menetelmissä kunkin kielen syntaksi eli lauseoppi ohjelmoidaan tarkasti, jolloin kullekin kielelle tarvitaan oma ohjelma-algoritmi. Eri kieliin liittyvien sanastojen tallennusta varten käytetään suurikapasiteettista, keski-

tettyä käänösmuistia. Esimerkkinä tällaisesta menetelmästä voidaan mainita Euroopan Unionin käänösjärjestelmä EuroTra. Tällaisilla ennestään tunnetuilla menetelmillä on useita epäkohtia. Syntaksin tarkka ohjelointi edellyttää erittäin suuren ohjelointityön. Tällainen syntaksialgoritmi samoin kuin tarvittava käänösmuistakin vaativat suuren muistilan tietokannassa. Koska näin toimiva käänösmenetelmä on monimutkainen, käänösten suorittaminen kohtuullisessa ajassa edellyttää tietokoneelta erittäin tehokasta suorituskykyä. Edellä mainittujen epäkohtien vuoksi käänökseen soveltuva laitteisto on kallis. Lisäksi tunnettuihin menetelmiin liittyvä nä 5 epäkohtana on se, että käänösalgoritmin päivitys edellyttää aina ohjelointityötä ja ohjelmapäivityksen tietokoneessa.

Esillä olevan keksinnön tarkoituksena on luoda ratkaisu informaation käänämiseksi, jonka avulla edellä esitetty, tekniikan tasoon liittyvät epäkohdat voidaan välttää.

10 15 Keksinnön eräänä oivalluksena on se, että käännettävä informaatio jaetaan rakenne- segmentteihin ja suoritetaan käänös rakennesegmenttikohtaisesti. Käänös suorite- taan tietämyskantaan tallennettujen mallisegmenttien ja sääntöjen perusteella. Tie- tämyskannan sisältämää tietoa laajennetaan edullisesti siten, että käänöksen suori- tuksen liittyen käyttäjältä kysytään tarvittaessa käyttöliitynnän välityksellä uusien 20 rakennesegmenttien käänöksiä, jotka sitten tallennetaan mallisegmentteinä tietä- myskantaan. Keksinnön mukaisen ratkaisun ansiosta käänöksiä suorittavassa lait- teistossa riittää pienempi muistikapasiteetti ja prosessorin nopeus. Lisäksi ohjel- mointityötä tarvitaan huomattavasti vähemmän ja laitteiston toimintaa voidaan ke- hittää ilman ohjelmistopäivityksiä.

25 30 Keksinnön mukaiselle menetelmälle ensimmäisen kielen merkkijonona esitetyn in- formaation käänämiseksi koneellisesti toisen kielen merkkijonoksi, jossa menetel- mässä

- tallennetaan tietämyskantaan mainittuina ensimmäisen kielen merkkijonoina olevia mallisegmenttejä ja niihin loogisesti liittyen toisen kielen merkkijonoina ole- via mallisegmenttejä,
- mainitusta ensimmäisen kielen merkkijonosta tunnistetaan rakennesegmentti ensimmäisen säännön perusteella,
- mainittua tunnistettua rakennesegmenttiä verrataan toisen säännön avulla tal- lennettuihin ensimmäisen kielen merkkijonoina oleviin mallisegmentteihin,
- mainitun vertailun perusteella pyritään valitsemaan yksi mallisegmentti,
- luetaan valittuun mallisegmenttiin loogisesti liittyvä toisen kielen merkkijono- na oleva mallisegmentti eli vastasegmentti ja

- mainittu rakennesegmentti käännetään mainituksi toisen kielen merkkijonona olevaksi käänössegmentiksi mainitun vastasegmentin ja kolmannen säännön perusteella,
- 5 on tunnusomaista se, että menetelmässä tunnistetaan väikesana ja/tai sanapääte ja mainittu ensimmäinen sääntö perustuu oleellisesti mainittuun väikesaan ja/tai sanapäätteen tunnistamiseen.

Keksinnön mukaiselle järjestelylle ensimmäisen kielen merkkijonona esitetyn informaation käänämiseksi toisen kielen merkkijonoksi, joka järjestely käsittää

- 10 - tietämyskantavälineet mainittuina ensimmäisen kielen merkkijonoina olevien mallisegmenttien ja niihin loogisesti liittyen toisen kielen merkkijonoina olevien vastasegmenttien tallentamiseksi sekä ensimmäisen, toisen ja kolmannen säännön tallentamiseksi,
- 15 - välineet rakennesegmentin tunnistamiseksi mainitusta ensimmäisen kielen merkkijonona esitetystä informaatiosta ensimmäisen säännön perusteella,
- välineet mainitun tunnistetun rakennesegmentin vertaamiseksi tallennettuihin ensimmäisen kielen merkkijonoina oleviin mallisegmentteihin toisen säännön perusteella,
- välineet yhden mallisegmentin valitsemiseksi mainitun vertailun perusteella,
- 20 - välineet valittuun mallisegmenttiin loogisesti liittyvän toisen kielen merkkijonona olevan mallisegmentin eli vastasegmentin lukemiseksi mainituista tietämyskantavälineistä ja
- välineet mainitun rakennesegmentin käänämiseksi mainituksi toisen kielen merkkijonona olevaksi käänössegmentiksi mainitun vastasegmentin ja kolmannen säännön perusteella, jolloin mainittu käänössegmentti edustaa mainitulla toisella kielellä esitettävää informaatiota,
- 25 on tunnusomaista se, että mainitut välineet rakennesegmentin tunnistamiseksi mainitusta ensimmäisen kielen merkkijonona esitetystä informaatiosta käsittävät välineet väikesanan ja/tai sanapäätteen tunnistamiseksi, jolloin mainittu ensimmäinen sääntö perustuu oleellisesti mainittuun väikesaan ja/tai sanapäätteen tunnistamiseen.

Keksinnön edullisia suoritusmuotoja on esitetty epäitsenäisissä patenttivaatimuksissa.

- 35 Keksintöä selostetaan seuraavassa tarkemmin oheisten piirustusten avulla, joissa
 - kuvio 1 esittää vuokaaviona erästä keksinnön mukaista menetelmää informaation käänämiseksi,

5 kuvio 2 esittää lohkokaaviona erästä keksinnön mukaista järjestelyä informaation
kääntämiseksi,

kuvio 3 esittää erästä rakennesegmentteihin jaettua teksti-informaatiota,

kuvio 4 esittää erään rakennesegmentin käännöstoimintoa, kun läheinen malliseg-
mentti esiintyy tietämyskannassa, ja

5 kuvio 5 esittää erään rakennesegmentin käännöstoimintoa, kun läheinen malliseg-
mentti ei esiinny tietämyskannassa.

10 Kuvio 1 esittää erästä keksinnön mukaista menetelmää informaation kääntämiseksi. Käännettävä informaatio ensin luetaan, lohko 101, ja jaetaan rakennesegmentteihin ensimmäisen säännön perusteella, lohko 102. Tämän jälkeen luetaan ensimmäinen rakennesegmentti kääntämättömästä informaatiosta, lohko 103. Luettua rakennesegmenttiä verrataan tietämyskantaan tallennettuihin mallisegmentteihin, lohkot 104 ja 110. Tällöin vertailu suoritetaan toisen säännön mukaisesti, jonka perusteella määräytyy, onko mallisegmentti käännettävään rakennesegmenttiin nähden läheinen. Jos tietämyskannasta löytyy kyseiseen rakennesegmenttiin nähden läheinen mallisegmentti, luetaan tietämyskannasta läheiseen mallisegmenttiin loogisesti liitetty, toisella kielellä oleva mallisegmentti eli vastasegmentti, lohko 121. Tämän jälkeen luetun vastasegmentin perusteella muodostetaan käännettävästä rakennesegmentistä 20 toiselle kielelle käännetty käännössegmentti kolmannen säännön mukaisesti, lohko 122. Tämän jälkeen tarkistetaan, onko jäljellä kääntämättömiä rakennesegmenttejä, lohko 123. Jos kääntämättömiä rakennesegmenttejä on jäljellä, palataan lohkoon 103, jossa luetaan seuraava kääntämätön rakennesegmentti käännöstä varten. Jos kääntämättömiä rakennesegmenttejä ei ole jäljellä lohkossa 123, järjestetään käännössegmentit lauseiksi neljännen säännön perusteella, minkä jälkeen käännetty informaatio tallennetaan. Tallennettu informaatio voidaan edelleen esittää esim. näytössä tai tulostaa esim. paperille tai levykkeelle, lohko 124.

30 Jos lohkossa 110 tietämyskannasta ei löydy rakennesegmentille läheistä mallisegmenttiä, esitetään ko. rakennesegmentti käyttöliityntävälilineellä, esim. näytöllä, lohko 131. Tämän jälkeen käyttäjä syöttää rakennesegmentin käännöksen eli vastasegmentin, lohko 132. Kyseiset rakennesegmentti ja vastasegmentti tallennetaan mallisegmenteiksi tietämyskantaan tulevaa käyttöä varten, lohkot 133, 134. Tämän jälkeen siirtytää lohkoon 123 ja jatketaan kuten edellä selostettiin. Tässä tapauksessa 35 vastasegmentti on yleensä suoraan käännössegmentti, jos käyttäjältä on kysytty rakennesegmentin käännöstä alkuperäisessä informaatiossa esitetyssä muodossa. Siten lohkon 122 toimintoa ei tässä tapauksessa välittämättä tarvita.

Mainittu ensimmäinen sääntö, jonka avulla rakennesegmentit tunnistetaan, voi perustua esim. ns. välikesanojen tai sijamuotojen tunnistukseen. Välikesanoja voivat olla esim. prepositiot ja partikkelit, jotka ovat yleensä vakioita merkkijonoja. Siten ne on mahdollista tunnistaa yksinkertaisesti vertaamalla kunkin sanan muodostavia merkkijonoja em. tunnettuihin välikesanien muodostaviin merkkijonoihin. Sijamuotojen tunnistus voidaan suorittaa esim. pääteiden eli suffiksien avulla vertailemalla sanojen viimeisiä merkkejä tunnettuihin pääteisiin. Sanan muodostavat merkkijonot voidaan erottaa toisistaan tunnetusti välimerkin avulla. Koska rakennesegmentti voi edullisesti käsittää useita sanoja, se voi sisältää myös yhden tai useamman välimerkin.

Mainittu toinen sääntö, jonka perusteella rakennesegmenttiä verrataan mallisegmentteihin, voi yksinkertaismillaan tarkoittaa yhtäläisyyttä. Tällöin siis tietämyskannasta etsitään tarkasti samaa mallisegmenttiä kuin esillä oleva käännettävä rakennesegmentti. Tietämyskannalta vaadittavan muistitilan kannalta on kuitenkin edullista, jos tietämyskantaan ei tallenneta esim. mallisegmentin eri sijamuotoja erikseen vaan toisen säännön avulla tunnistetaan myös eri sijamuodossa oleva mallisegmentti. Tällöin mallisegmenttiin loogisesti liittyvä vastasegmentti on myös muuttava tarvittavaan sijamuotoon käänössegmentin muodostamiseksi. Tämä tapahtuu kolmannen säännön perusteella, joka siis tällöin käsittää tiedot ko. kielen sijamuodoista.

Mainittu neljäs sääntö, jonka perusteella käänössegmentit järjestetään käännytyksi lauseiksi, tarkoittaa useissa tapauksissa käänössegmenttien asettamista samaan järjestykseen, jossa käännettävät rakennesegmentit olivat ensimmäisessä kielessä. Järjestys voi kuitenkin olla kiestä riippuva, joten myös mainittu neljäs sääntö on kielikohtainen.

Mallisegmentin tallennuksen yhteydessä voidaan edullisesti tallentaa myös mallisegmentin lajitunnus. Tällöin lajitunnus tallennetaan kuhunkin mallisegmenttiin loogisesti liittyvänä. Jos lajitunnuksia käytetään, voidaan rakennesegmentin tunnistukseen ja käänökseen mallisegmentin perusteella soveltaa eri sääntöjä riippuen rakennesegmentin lajista. Rakennesegmenttien lajeja voivat olla mm. teon kohde, erisnimi, verbi, paikan ilmaisu, adjektiivi ja idiomi. Jos lajitunnuksia käytetään, kysytään käyttäjältä rakennesegmentin käänöksen yhteydessä myös, mihin lajiin ko. rakennesegmentti ja sen käänös kuuluvat.

Eräs keksintöön liittyvä oivallus on se, että tietämyskantaa päivitetään interaktiivisesti suoritettavan käänöstapahtuman yhteydessä. On huomattava, että tietämyskannan päivityksen ei tarvitse rajoittua uusien malli- ja vastasegmenttien tallentamiseen, vaan myös mainittuja sääntöjä voidaan edullisesti päivittää. Tällöin päivitys 5 tapahtuu esim. käyttäjän syöttämän uuden rakennesegmentin käänökseen yhteydessä tunnistamalla syötettyyn käänökseen liittyvää säännönmukaisuus.

Edellä on esitetty yhden informaation käänämistä ensimmäiseltä kieleltä toiselle kielelle. Tietämyskantaan suoritettuja päivityksiä käytetään edullisesti hyväksi seuraavien informaatioiden käänämisessä. Siten keksinnön mukainen menettely toisistaan seuraavien ensimmäisen ja toisen informaation käänämiseksi voi koostua esim. seuraavista vaiheista:

- luetaan ensimmäisen kielen merkkijonona esitetyt ensimmäinen informaatio,
- suoritetaan tietämyskannassa olevien tietojen perusteella mainitun ensimmäisen kielen merkkijonona esitetyn ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi siltä osin kuin se on tietämyskannassa olevien tietojen perusteella mahdollista,
- määritetään ne lisätiedot, jotka tarvitaan ensimmäisen kielen merkkijonona esitetyn ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi loppuun suorittamiseksi,
- syötetään tietämyskantaan mainitut lisätiedot tietämyskannan päivittämiseksi,
- suoritetaan loppuun ensimmäisen kielen merkkijonona esitetyn ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi,
- tallennetaan mainittu toisen kielen merkkijonona esitetyt ensimmäinen informaatio,
- tallennetaan ensimmäisen kielen merkkijonona esitetyt toinen informaatio,
- suoritetaan tietämyskannassa olevien mainittujen päivitettyjen tietojen perusteella mainitun ensimmäisen kielen merkkijonona esitetyn toisen informaation käänäminen toisen kielen merkkijonona esitetyksi toiseksi informaatioksi.

Kuvio 2 esittää lohkokaaviota eräästä keksinnön mukaisesta laitejärjestelystä informaation käänämiseksi. Siinä prosessoriin 20 liittyvät käyttöliityntäväliset levyke-35 asema 21, näyttö 22 ja näppäimistö 23. Levykeaseman avulla laitteeseen voidaan syöttää levykkeeltä käännettävä informaatio ja tallentaa käännetty informaatio levykkeelle informaation käänämiseksi muissa laitteissa. Ko. informaatiota voidaan siirtää laitteen ja muiden tiedonkäsittelylaitteiden välillä myös tiedonsiirtoväylän I/O

kautta. Näyttöä 22 voidaan käyttää sellaisten rakennesegmenttien esittämiseen käyttäjälle, joiden käänöstä ei löydy tietämyskannasta. Käyttäjä voi syöttää tällaisen rakennesegmentin käänöksen näppäimistön 23 avulla. Em. käyttöliityntävälaineitä voidaan käyttää myös käännetyn informaation tarkastuksessa ja korjauksessa.

5

Kuvion 2 mukainen laite käsittää lisäksi sähköisen muistin 24 mm. käännettävien rakennesegmenttien ja käänössegmenttien väliaikaista tallennusta varten. Lisäksi laitteessa on massamuisti 25 mm. tietämyskannan eli mallisegmenttien, lajittunusten ja sääntöjen sekä ohjelmien tallennusta varten. Massamuistina voidaan käyttää esim. kovalevyasemaa tai optista levykeasemaa. Edellä mainitut laiteosat voidaan toteuttaa ennestään tunnettujen tietokonekomponenttien avulla saattamalla ne erityisen ohjelmiston avulla toimimaan keksinnön mukaisesti. Laiteosien välillä merkkijonot ym. tiedot siirretään edullisesti sähköisinä signaaleina.

10

15 Keksinnön soveltaminen ei millään tavoin rajoitu edellä esitettyihin komponentteihin vaan keksinnön mukainen järjestely voidaan toteuttaa lukuisilla tavoilla, jotka alan ammattilainen pystyy tämän selostuksen perusteella suunnittelemaan.

20

Kuvio 3 esittää erästä rakennesegmentteihin 31, 32, 33 ja 34 jaettua englanninkielistä lausetta. Kuten huomataan, rakennesegmentti käsittää tyypillisesti lauseessa toisiaan seuraavia sanoja, jotka liittyvät läheisesti toisiinsa. Siten rakennesegmentti sisältää usein myös sanoja erottavan välimerkin.

25

Kuvio 4 esittää kuviossa 3 esitetyn lauseen ensimmäisen rakennesegmentin käänöstä erään keksinnön mukaisen ratkaisun avulla. Siinä käännettävä rakennesegmentti 42 on tallennettu käänösmuistiin 41 ja ko. rakennesegmenttiä verrataan tietämyskantaan 44 tallennettuihin mallisegmentteihin. Kuvion 4 esittämässä tapauksessa kyseinen rakennesegmentti on jo aiemmin tallennettu tietämyskantaan mallisegmentiksi 45, joka vertailussa löydetään. Jos esim. halutaan käääntää esillä oleva informaatio suomeksi, luetaan tietämyskannasta edellä mainittuun englanninkieliseen mallisegmenttiin loogisesti liitetty suomenkielinen mallisegmentti 46. Kuviossa 4 mallisegmenttien 45 ja 46 välillä oleva kaksoisviiva kuvailee loogista liittymistä. Kun suomenkielinen mallisegmentti on luettu, se tallennetaan käänössegmentiksi käänösmuistiin.

30

35 Kuvio 5 esittää kuviossa 3 esitetyn lauseen toisen rakennesegmentin käänöstä erään keksinnön mukaisen ratkaisun avulla. Tässä tapauksessa tietämyskantaan ei ole ennestään tallennettu mallisegmentiksi käännettävä englanninkielistä rakenne-

segmenttiä tai sen suomenkielistä vastasegmenttiä tai kumpaakaan niistä. Tässä tapauksessa käänösmuistiin 51 tallennettua käännettävää rakennesegmenttiä 52 vertataan tietämiskannan mallisegmentteihin, ja kun tietämiskannasta ei löydetä haluttua vastasegmenttiä, esitetään käännettävä rakennesegmentti 58 käyttöliittymän 57 näytöllä. Tämän jälkeen käyttäjä syöttää rakennesegmentin 58 käänöksen 59 käyttöliittymän välityksellä tietämiskantaan 54. Siten tietämiskantaan tallennetaan toisiinsa loogisesti liittyen englanninkielinen ja suomenkielinen mallisegmentti. Tämän jälkeen rakennesegmentin suomenkielinen käänös tallennetaan käänössegmentiksi 53 käänösmuistiin 51.

10 Jos tämän jälkeen edellä mainitut rakennesegmentit esiintyvät uudelleen käännettävässä informaatiossa, niitä vastaavat malli- ja vastasegmentit löytyvät tietämiskannasta, joten niitä ei tarvitse kysyä käyttäjältä uudelleen. Mutta jos seuraavaksi käännettävässä informaatiossa esiintyy lause "we have expanded our operation largely in Finland", esiintyy tässä uutena rakennesegmenttinä "largely". Jos sille läheistä mallisegmenttiä ei ennestään ole tallennettu tietämiskantaan, kysytään sen käänös käyttäjältä ja tallennetaan "largely" mallisegmenttinä tietämiskantaan ja siihen loogisesti liitettynä myös sen käänös, jonka käyttäjä on syöttänyt.

15 20 On huomattava, että laitteiston toiminta voidaan järjestää siten, että käänöstyö suoritetaan ensin koneellisesti koko informaatiolle siltä osin kuin se on tietämiskantaan tallennettujen mallisegmenttien perusteella mahdollista. Tämän jälkeen käyttäjä voi syöttää tarvittavat uusien rakennesegmentien käänökset tietämiskantaan. Tällaisen järjestelyn etuna on se, että käyttäjän ei tarvitse odottaa laitteen ääressä käänöstyön suoritusta vaan voi suorittaa tietämiskannan täydentämisen kerralla hänelle sopivana ajankohtana.

25 30 35 Mallisegmentit voidaan tallentaa tietämiskantaan segmenttipareina, jolloin kutakin kieliparia varten tallennetaan omat mallisegmenttiparit. Toinen mahdollisuus on liittää loogisesti toisiinsa useamman kielen mukaisia mallisegmenttejä, jolloin samoja mallisegmenttejä voidaan käyttää sellaisenaan useiden kieliparien käänösten yhteydessä. Tällöin kunkin kielen mallisegmenttejä voidaan täydentää tietämiskantaan aina silloin, kun ne esiintyvät kyseisellä kielellä ensimmäisen kerran. Kun silloin tietoa syötetään tietämiskantaan yhden kieliparin käänöksen yhteydessä, lisääntyy tietämiskannan sisältämä informaatio automaattisesti myös muissa kielipareissa.

Keksinnön mukainen ratkaisu ei periaatteeltaan ole kielisidonnainen, vaan sitä voidaan soveltaa mille tahansa kieliparille. Keksinnön käyttö ei myöskään ole rajoitettu tavanomaisessa kommunikaatiossa käytettäviin ns. luonnollisiin kieliin vaan sitä voidaan käyttää minkä tahansa merkkijonoista koostuvan kielen käänämiseksi toiseksi merkkijonoista koostuvalle kielelle. Esimerkinä tällaisista muista kielistä voidaan mainita ohjelointikielet ja tiedonsiirtoprotokollat.

10 Keksinnön mukaisella ratkaisulla on useita etuja ennestään tunnettuun tekniikan tasoon verrattuna. Ratkaisun toiminta edellyttää vain vähäisen kielikohtaisen tiedon ko. kielen jakamiseksi rakennesegmenteiksi. Toinen ratkaisun etu on siinä, että suorituksen aikana muistiin kerätään lisää tietoa, jolloin laite "oppii" uusia mallisegmenttipareja ja sääntöjä. Siten yksinkertaisella laitekokoonpanolla ja vähäisellä ohjelointi- ja päivitystyöllä voidaan aikaansaada tehokas väline koneellisten käänösten suorittamiseksi.

15 Keksinnön mukainen ratkaisu soveltuu hyvin tilanteisiin, joissa eksinnön mukaista järjestelyä käytetään usean käyttäjän tarpeisiin. Tällöin järjestely käsittää edullisesti useita käyttöliityntöjä, jotka voivat olla yhteydessä tietämyskantaan esim. tiedonsiirtoverkon välityksellä. Tietämyskanta voi tällöin olla edullisesti hajautettu siten, että ensimmäistä tietämyskantaa eli päätietämyskantaa voivat käyttää määrätyt käyttäjäjoukko ja toista tietämyskantaa eli alitietämyskantaa voi käyttää vain osa tällaisesta määräystä käyttäjäjoukosta. Näin eri käyttäjät voivat päivittää esim. erikoissanastolla tai -sanonnoilla omaa tietämyskantaansa, eikä tällaista erikoiskäyttöön soveltuva tietoa tällöin käytetä muiden käyttäjien toimesta.

20 25 Tällaisessa hajautetussa tietämyskannassa voidaan suorittaa ensimmäisen eli päätietämyskannan päivitys toisista eli alitietämyskannoista. Tällöin siirretään toisiin tietämyskantoihin tallennettua tietoa ensimmäiseen tietämyskantaan ennalta määrätyjen kriteerien perusteella. Tällainen kriteeri voi olla esim. se, kuinka usein tiettyä tietoa on käytetty. Tiedonsiirto tietämyskantojen välillä voidaan suorittaa myös siten, että yhteisen päätietämyskannan ylläpitäjä tarkastaa ja hyväksyy jokaisen siirrettävän tiedon.

30 35 Edellä on esitetty eräitä eksinnön mukaisen ratkaisun suoritusmuotoja. Luonnollisesti eksinnön mukaista periaatetta voidaan muunnella patenttivaatimusten suojaalueen puitteissa esim. toteutuksen yksityiskohtien sekä käyttöalueiden osalta.

Patenttivaatimuksset

1. Menetelmä ensimmäisen kielen merkkijonona esitetyn informaation käänämiseksi koneellisesti toisen kielen merkkijonoksi, jossa menetelmässä
 - 5 - tallennetaan tietämyskantaan mainittuina ensimmäisen kielen merkkijonoina olevia mallisegmenttejä ja niihin loogisesti liittyen toisen kielen merkkijonoina olevia mallisegmenttejä (133, 134),
 - mainitusta ensimmäisen kielen merkkijonosta tunnistetaan rakennesegmentti ensimmäisen säännön perusteella (102),
- 10 - mainittua tunnistettua rakennesegmenttiä verrataan toisen säännön avulla tallennettuihin ensimmäisen kielen merkkijonoina oleviin mallisegmentteihin (104),
 - mainitun vertailun perusteella pyritään valitsemaan yksi mallisegmentti (110),
 - luetaan valittuun mallisegmenttiin loogisesti liittyvä toisen kielen merkkijono oleva mallisegmentti eli vastasegmentti (121) ja
- 15 - mainittu rakennesegmentti käännetään mainituksi toisen kielen merkkijonona olevaksi käänössegmentiksi mainitun vastasegmentin ja kolmannen säännön perusteella (122),
tunnettu siitä, että menetelmässä tunnistetaan välikesana ja/tai sanapääte ja mainittu ensimmäinen sääntö perustuu oleellisesti mainittuun välikesaan ja/tai sanapäätteen tunnistamiseen.
- 20
- 25 2. Patenttivaatimuksen 1 mukainen menetelmä, tunnettu siitä, että mainittu toisen kielen merkkijonona esitettävä informaatio muodostetaan käänössegmenttien ja neljännen säännön perusteella (124).
- 30 3. Patenttivaatimuksen 1 tai 2 mukainen menetelmä, tunnettu siitä, että milloin rakennesegmenttien vertailun tuloksena ei löydy toisen säännön perusteella valittavaa mallisegmenttiä, rakennesegmentti esitetään käyttöliitynnän avulla (131) ja esitetyn rakennesegmentin vastasegmentti tallennetaan tietämyskantaan käyttöliitynnän avulla (132, 133).
- 35 4. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että mainittu rakennesegmentti sisältää välimerkin.
5. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että mallisegmenttiin loogisesti liittyvästi tallennetaan sen lajitunnus.

6. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että toisiinsa loogisesti liitettyjä, eri kieliä edustavia mallisegmenttejä on enemmän kuin kaksi.

5 7. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että jotakin mainittua sääntöä päivitetään käyttöliitynnästä syötetyn tiedon perusteella.

10 8. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että käyttöliitynnän avulla syötetään informaatiota tietämyskannan päivittämiseksi ensimmäisen informaation käänämiseksi ja mainittua syötettyä tietoa käytetään muun kuin mainitun ensimmäisen informaation käänämiseen tarvittavan tiedon päivittämiseen mainitussa tietämyskannassa.

15 9. Jonkin edellisen patenttivaatimuksen mukainen menetelmä, tunnettu siitä, että se käsittää vaiheet, joissa

- luetaan ensimmäisen kielen merkkijonona esitetyt ensimmäinen informaatio,
- suoritetaan tietämyskannassa olevien tietojen perusteella mainitun ensimmäisen kielen merkkijonona esitetyt ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi siltä osin kuin se on tietämyskannassa olevien tietojen perusteella mahdollista,
- määritetään ne lisätiedot, jotka tarvitaan ensimmäisen kielen merkkijonona esitetyt ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi loppuun suorittamiseksi,
- syötetään tietämyskantaan mainitut lisätiedot tietämyskannan päivittämiseksi,
- suoritetaan loppuun ensimmäisen kielen merkkijonona esitetyt ensimmäisen informaation käänäminen toisen kielen merkkijonona esitetyksi ensimmäiseksi informaatioksi,
- tallennetaan mainittu toisella kielellä esitetyt ensimmäinen informaatio,
- luetaan ensimmäisen kielen merkkijonona esitetyt toinen informaatio,

20 30 10. Järjestely ensimmäisen kielen merkkijonona esitetyt informaation käänäminiseksi toisen kielen merkkijonoksi, joka järjestely käsittää

- tietämyskantaväliset (20, 25) mainittuna ensimmäisen kielen merkkijonoina olevien mallisegmenttien ja niihin loogisesti liittyen toisen kielen merkkijonoina

olevien vastasegmenttien tallentamiseksi sekä ensimmäisen, toisen ja kolmannen säännön tallentamiseksi,

- välineet (20, 24) rakennesegmentin tunnistamiseksi mainitusta ensimmäisen kielen merkkijonona esitetystä informaatiosta ensimmäisen säännön perusteella,
- 5 - välineet (20, 25) mainitun tunnistetun rakennesegmentin vertaamiseksi tallennettuihin ensimmäisen kielen merkkijonoina oleviin mallisegmentteihin toisen säännön perusteella,
- välineet (20) yhden mallisegmentin valitsemiseksi mainitun vertailun perusteella,
- 10 - välineet (20, 25) valittuun mallisegmenttiin loogisesti liittyvän toisen kielen merkkijonona olevan mallisegmentin eli vastasegmentin lukemiseksi mainitusta tietämyskantavälaineistä ja
- välineet (20, 24) mainitun rakennesegmentin käänämiseksi mainituksi toisen kielen merkkijonona olevaksi käänössegmentiksi mainitun vastasegmentin ja kolmannen säännön perusteella, jolloin mainittu käänössegmentti edustaa mainitulla 15 toisella kielellä esitettävää informaatiota,
- tunnettu siitä, että mainitut välineet (20, 24) rakennesegmentin tunnistamiseksi mainitusta ensimmäisen kielen merkkijonona esitetystä informaatiosta käsittevät välineet välikesan ja/tai sanapäätteen tunnistamiseksi, jolloin mainittu ensimmäinen 20 sääntö perustuu oleellisesti mainittuun välikesan ja/tai sanapäätteen tunnistamiseen.
- 11. Patenttivaatimuksen 10 mukainen järjestely, tunnettu siitä, että se lisäksi käsitteää välineet (20, 25) toisen kielen merkkijonona esitettävän informaation muodostamiseksi vähintään kahden käänössegmentin ja neljännen säännön perusteella.
- 25 12. Patenttivaatimuksen 10 tai 11 mukainen järjestely, tunnettu siitä, että se käsitteää käyttöliityntävälineet (22, 23) käyttäjän liittymiseksi mainittuihin tietämyskantavälaineisiin.
- 30 13. Patenttivaatimuksen 12 mukainen järjestely, tunnettu siitä, että mainitut käyttöliityntävälineet liittyvät mainittuihin tietämyskantavälaineisiin tiedonsiirtoverkon välityksellä.
- 35 14. Jonkin patenttivaatimuksen 10-13 mukainen järjestely, tunnettu siitä, että mainitut tietämyskantavälaineet käsittevät ensimmäisen tietämyskantavälaineen (25) ja toisen tietämyskantavälaineen siten, että mainittuun ensimmäiseen tietämyskantaväli-

neeseen on pääsy määrätyillä käyttäjillä ja mainittuun toiseen tietämyskantavälineeseen on pääsy vain osalla mainituista määrätyistä käyttäjistä.

5 15. Jonkin patenttivaatimuksen 10-14 mukainen järjestely, tunnettu siitä, että mainittut tietämyskantavälineet käsittävät ensimmäisen tietämyskantavälineen (25) ja toisen tietämyskantavälineen, jolloin järjestely käsittää välineet tiedon syöttämiseksi käyttöliityntävälineistä mainittuun toiseen tietämyskantavälineeseen sekä välineet mainittuun toiseen tietämyskantaan tallennetun tiedon siirtämiseksi valikoidusti mainittuun ensimmäiseen tietämyskantavälineeseen.

10 Patentkrav

1. Förfarande för mekanisk översättning av information given som en teckensträng på ett första språk till en teckensträng på ett andra språk, varvid
 - i kunskapsbasen lagras modellsegment som teckensträngar på det första språket och, i logisk anslutning till dessa modellsegment (133, 134), som teckensträngar på det andra språket,
 - i nämnda teckensträng på det första språket identifieras ett struktursegment utgående från en första regel (102),
 - nämnda identifierade struktursegment jämförs med modellsegment (104) som med en andra regel lagrats som teckensträngar på det första språket,
 - utgående från nämnda jämförelse strävar man efter att välja ett modellsegment (110),
 - avläses ett modellsegment, dvs. motsegment (121) som en teckensträng på det andra språket i logisk anslutning till det valda modellsegmentet och
 - nämnda struktursegment översätts till nämnda översättningssegment som en teckensträng på det andra språket utgående från nämnda motsegment och en tredje regel (122),
 kännetecknat av att vid förfarandet identifieras ett mellanord och/eller ändelse och nämnda första regel baserar sig väsentligen på identifieringen av nämnda mellanord och/eller ändelse.
2. Förfarande enligt patentkrav 1, kännetecknat av att informationen som ges som en teckensträng på det andra språket genereras utgående från översättningssegmenten och en fjärde regel (124).
3. Förfarande enligt patentkrav 1 eller 2, kännetecknat av att då jämförelsen av struktursegmenten inte resulterar i ett modellsegment att väljas enligt den andra re-

geln, visas struktursegmentet med ett användargränssnitt (131) och motsegmentet till det visade struktursegmentet lagras i kunskapsbasen med hjälp av användargränsnittet (132, 133).

5 4. Förfarande enligt något av föregående patentkrav, kännetecknat av att nämnda struktursegment innehåller ett mellantecken.

5. Förfarande enligt något av föregående patentkrav, kännetecknat av att i logisk anslutning till modellsegmentet lagras dess artidentifierare.

10 6. Förfarande enligt något av föregående patentkrav, kännetecknat av att det finns flera än två logiskt förbundna modellsegment som företräder olika språk.

15 7. Förfarande enligt något av föregående patentkrav, kännetecknat av att något av nämnda regler uppdateras utgående från information som inmatats i användargränsnittet.

20 8. Förfarande enligt något av föregående patentkrav, kännetecknat av att med gränsnittet inmatas information för uppdatering av kunskapsbasen för översättning av den första informationen och nämnda inmatade information används för uppdatering av information som behövs för översättning av någon annan information än nämnda första information i nämnda kunskapsbas.

25 9. Förfarande enligt något av föregående patentkrav, kännetecknat av att det innehåller steg för att

- läsa den första informationen given som en teckensträng på det första språket,
- utgående från data i kunskapsbasen utföra översättning av nämnda första information i form av en teckensträng på det första språket till första information i form av en teckensträng på det andra språket i den mån det är möjligt på basen av data som finns i kunskapsbasen,
- bestämma de tilläggsdata som behövs för slutförandet av översättningen av den första informationen i form av en teckensträng på det första språket till en första information i form av en teckensträng på det andra språket,
- i kunskapsbasen inmata nämnda tilläggsdata för uppdatering av kunskapsbasen,
- slutföra översättningen av den första informationen given som en teckensträng på det första språket till en första information given som en teckensträng på det andra språket,

- lagra nämnda första information given på det andra språket,
- läsa den andra informationen given som en teckensträng på det första språket,
- utgående från nämnda uppdaterade data i kunskapsbasen utföra översättning av den andra informationen given som en teckensträng på det första språket till andra information given som en teckensträng på det andra språket.

5 10. Arrangemang för översättning av information given som en teckensträng på ett första språk till en teckensträng på ett andra språk, innehållande

- kunskapsbasorgan (20, 25) för att lagra modellsegment i form av nämnda teckensträngar på det första språket och i logisk anslutning till dessa motsegment i form av teckensträngar på det andra språket samt för att lagra den första, andra och tredje regeln,
- organ (20, 24) för att identifiera struktursegment i informationen som givits som en teckensträng på det första språket enligt den första regeln,

10 15 20 25 30 35 15. organ (20, 25) för att jämföra nämnda identifierade struktursegment med de lagrade modellsegmenten i form av teckensträngar på det första språket enligt den andra regeln,

- organ (20) för att välja ett modellsegment på grundval av nämnda jämförelse,
- organ (20, 25) för att läsa modellsegmentet, dvs. motsegmentet i form av en teckensträng på det andra språket i logisk anslutning till det valda modellsegmentet i nämnda kunskapsbasorgan och
- organ (20, 24) för att översätta nämnda struktursegment till nämnda översättningssegment i form av en teckensträng på det andra språket enligt nämnda motsegment och den tredje regeln, varvid nämnda översättningssegment företräder informationen som skall ges på nämnda andra språk,

kännetecknat av att nämnda organ (20, 24) för identifiering av struktursegmentet i informationen given som en teckensträng på det första språket innehåller organ för identifiering av ett mellanord och/eller en ändelse, varvid nämnda första regel baseras sig väsentligen på identifieringen av nämnda mellanord och/eller ändelse.

11. Arrangemang enligt patentkrav 10, kännetecknat av att det vidare innehåller organ (20, 25) för att generera information att ges som en teckensträng på det andra språket enligt minst två översättningssegment och den fjärde regeln.

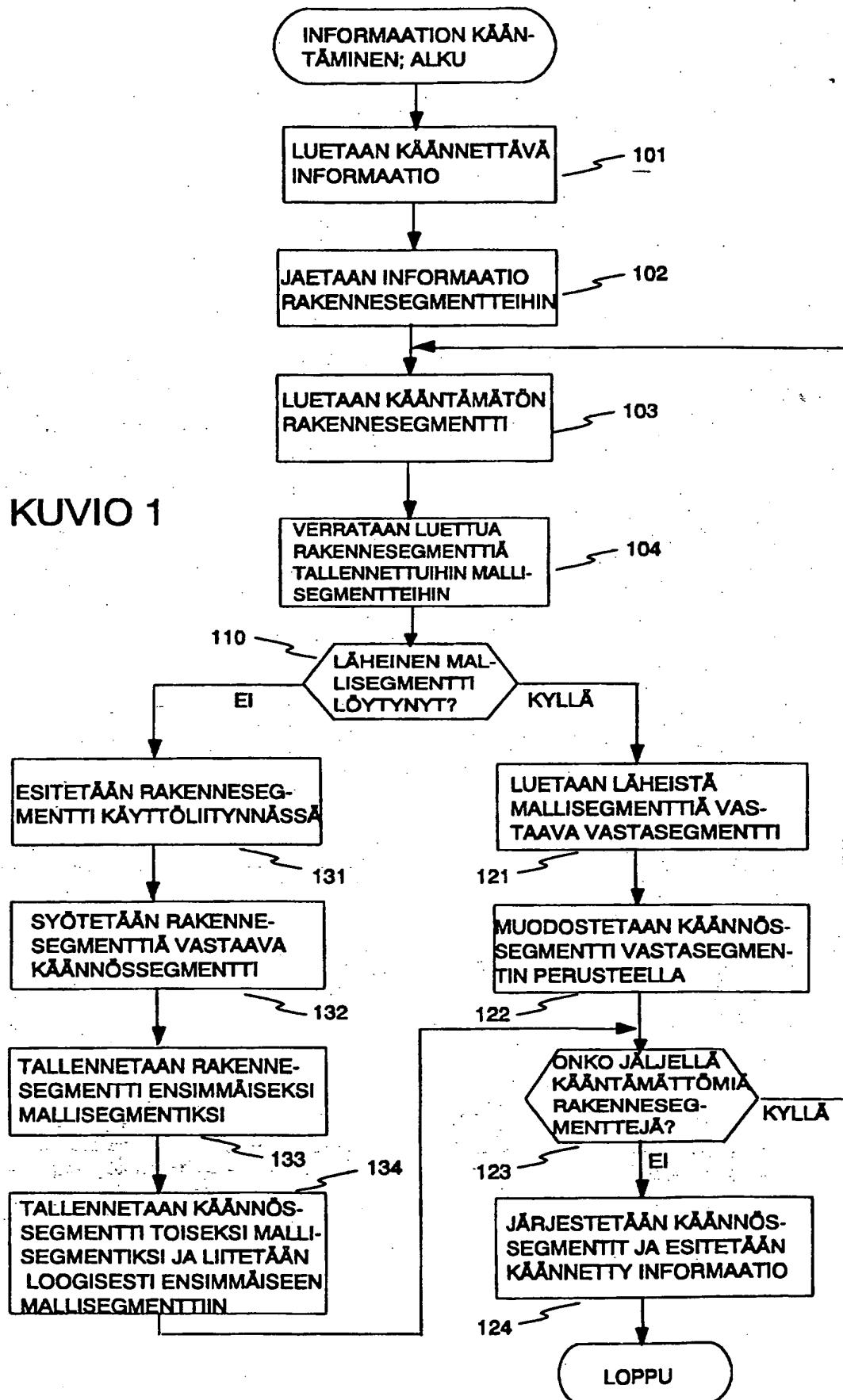
12. Arrangemang enligt patentkrav 10 eller 11, kännetecknat av att det innehåller användargränssnittsorgan (22, 23) för anslutning av användaren till nämnda kunskapsbasorgan.

13. Arrangemang enligt patentkrav 12, kännetecknat av att nämnda användargränssnittsorgan är anslutna till nämnda kunskapsbasorgan via ett dataöverföringsnät.

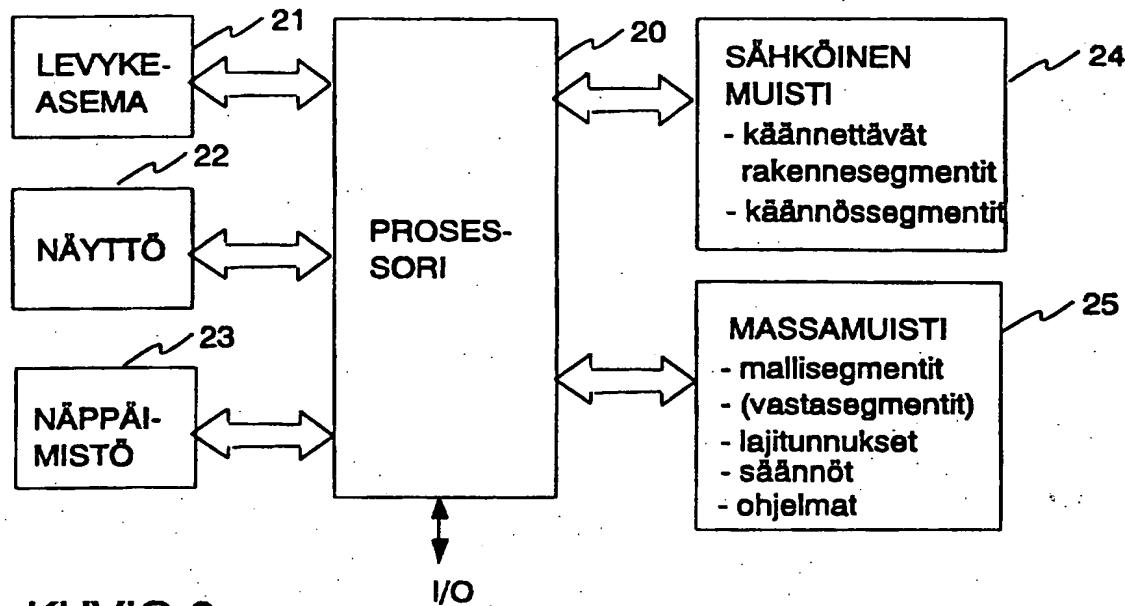
5 14. Arrangemang enligt något av patentkraven 10-13, kännetecknat av att nämnda kunskapsbasorgan innehåller ett första kunskapsbasorgan (25) och ett andra kunskapsbasorgan så att vissa användare har åtkomst till nämnda första kunskapsbasorgan och endast en del av nämnda vissa användare har åtkomst till nämnda andra kunskapsbasorgan.

10 15. Arrangemang enligt något av patentkraven 10-14, kännetecknat av att nämnda kunskapsbasorgan innehåller ett första kunskapsbasorgan (25) och ett andra kunskapsbasorgan, varvid arrangemanget innehåller organ för inmatning av data från användargränssnittsorganen i nämnda andra kunskapsbasorgan samt organ för selektiv överföring av data lagrade i nämnda andra kunskapsbas i nämnda första kunskapsbasorgan.

KUVIO 1



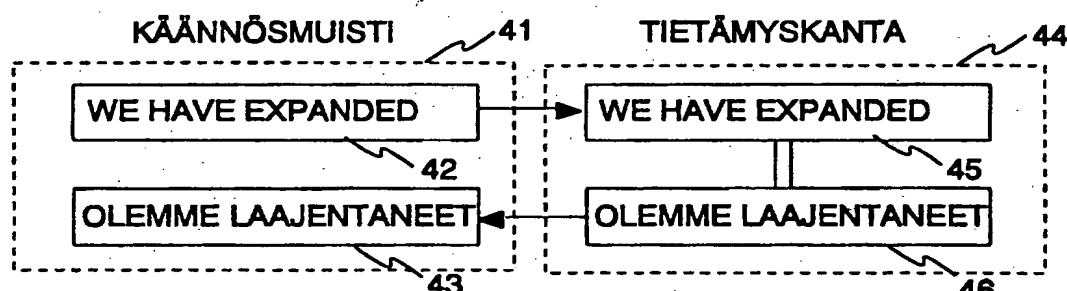
This Page Blank (uspto)



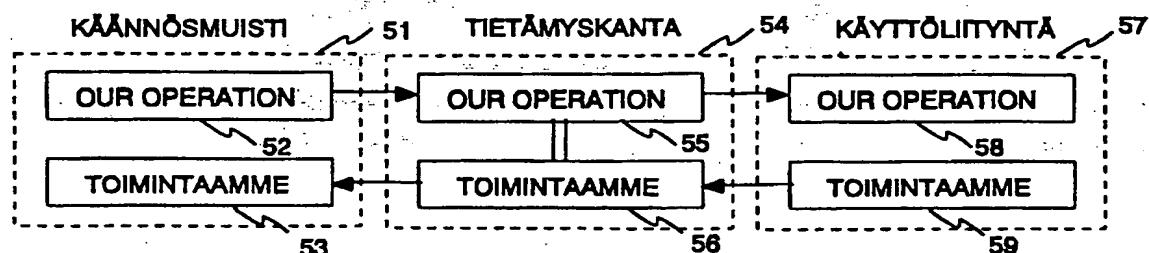
KUVIO 2

WE HAVE EXPANDED 31 OUR OPERATION 32 IN FINLAND 33 CONSIDERABLY 34

KUVIO 3



KUVIO 4



KUVIO 5

This Page Blank (usptc)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)